Change Detection Techniques for Dynamic Network Data

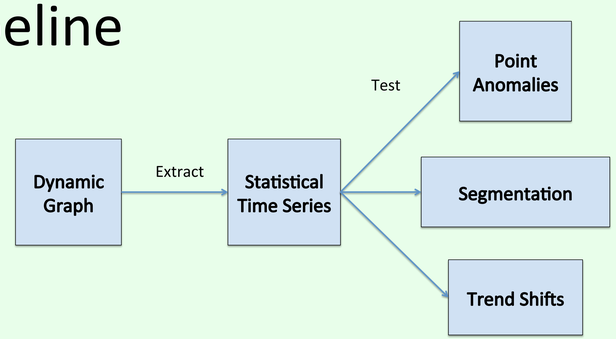
Timothy La Fond[12], Brian Gallagher[2], Jennifer Neville[1]

Purdue University[1], Lawrence Livermore National Labs[2]

Introduction

Automatic detection of changes or anomalies in time series data is an important topic in data analysis. However, when the time series is composed of network snapshots the task becomes exponentially more difficult. Examples of these dynamic networks are the activity levels of a computer network over 24 hours, or the e-mail communications of a corporation over a month. Improved change detection for dynamic networks allows the detection of problems or intrusions in these networks.

Pipeline

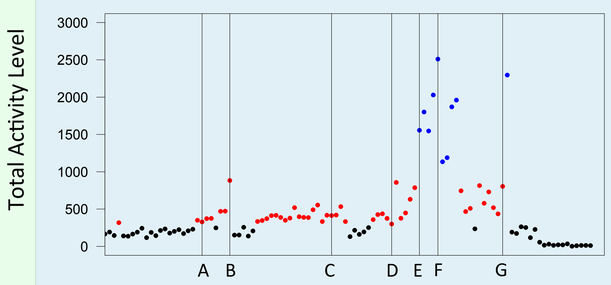


- Dynamic graph very complex
- Reduce to multiple time series using graph statistics
- Point Anomaly: single unusual time step
- Segmentation: find distinct regions in time
- Trend Shifts: graph behavior changes

Detecting Major Enron Events

A: (11/1/2000) FERC report on California energy crisis
B: (12/13/2000) Announcement that Skilling will become CEO
C: (5/17/2001) Secret meeting with Governor Schwarzenegger
D: (8/14/2001) Skilling approaches Lay about resigning

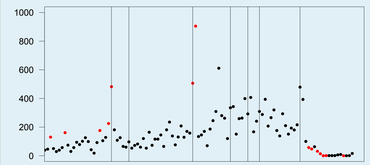
E: (9/26/2001) Lay tells employees Enron stock is a bargain
F: (10/24/2001) Enron stock drops by 50%
G: (2/6/2002) Skilling testifies before Congress



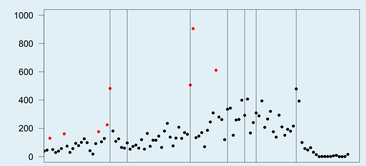
Create regions by k-means clustering values.
Regions roughly correspond to Enron time periods.
Linear segmentation is alternative approach.

Top 5 maximum points for each statistic.
Detected points often align with key Enron events.

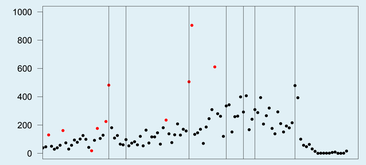
Online Change Detection



Empirical Reference Distribution



Gaussian Reference Distribution



Linear Fit with Residual Distribution

- Flag events as anomalous as they occur, using only previous time steps
- Construct reference distribution using 50 previous time steps
- Use statistical test to determine if current is anomalous
- Hindered by less available information
- Empirical: collect reference distribution and set confidence interval on observations
- Gaussian: model observations as normally distributed, reject points based on probability
- Linear Fit: use linear regression line, distance to line becomes abnormality measure

Future Work

- Examples of network statistics: which one is the best?
- Aggregate statistics for better conclusion
- Trend Shifts: so far, can find similar regions or unusual points in time. Would want to determine when behavior changes, i.e. slope of graph shifts
- Model-based approach: difficult to represent complex network behavior with model, but could be comprehensive approach

Conclusions

This work has described the problem of change detection in dynamic network data and provided the framework for performing such detections. E-mail data from the Enron corporation provides an example of detecting important network events using a variety of statistics. Finally, an example of how to detect events as they occur has been proposed and the effectiveness of this technique examined. In the future superior network change detection techniques will be a powerful tool for network analysts and cyber defense personnel.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.
LLNL-POST-641906